Journal of Experimental Psychology: Human Perception and Performance

Decoupling Object Detection and Categorization

Michael L. Mack, and Thomas J. Palmeri

Online First Publication, August 23, 2010. doi: 10.1037/a0020254

CITATION

Mack, M. L., & Palmeri, T. J. (2010, August 23). Decoupling Object Detection and Categorization. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0020254

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/10/$12.00 DOI: 10.1037/a0020254

Decoupling Object Detection and Categorization

Michael L. Mack and Thomas J. Palmeri
Vanderbilt University

We investigated whether there exists a behavioral dependency between object detection and categorization. Previous work (Grill-Spector & Kanwisher, 2005) suggests that object detection and basic-level categorization may be the very same perceptual mechanism: As objects are parsed from the background they are categorized at the basic level. In the current study, we decouple object detection from categorization by manipulating the between-category contrast of the categorization decision. With a superordinate-level contrast with people as one of the target categories (e.g., cars vs. people), which replicates Grill-Spector and Kanwisher, we found that success at object detection depended on success at basic-level categorization and vice versa. But with a basic-level contrast (e.g., cars vs. boats) or superordinate-level contrast without people as a target category (e.g., dog vs. boat), success at object detection did not depend on success at basic-level categorization. Successful object detection could occur without successful basic-level categorization. Object detection and basic-level categorization do not seem to occur within the same early stage of visual processing.

Keywords: recognition, categorization, detection, basic level, time course

When is an object's category known during visual perception? Typically, objects are categorized faster at the so-called basic level (e.g., dog or car) than more subordinate (e.g., Doberman Pinscher or Honda Civic) or superordinate (e.g., animal or vehicle) levels (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), although the relative timing may be malleable (Jolicoeur, Gluck, & Kosslyn, 1984; Tanaka & Taylor, 1991; Tanaka, 2001). The basic level could be fastest because basic-level categorization must precede finer or coarser levels of categorization during visual processing (Jolicoeur et al., 1984)—not only are basic-level categorizations fast, but basic-level categorization takes place first (but see Mack, Wong, Gauthier, Tanaka, & Palmeri, 2007, 2009; Palmeri, Wong, & Gauthier, 2004). Grill-Spector and Kanwisher (2005) specifically placed basic-level categorization at a relatively early stage of visual processing, suggesting that basic-level categorization could be intimately tied to figure-ground segmentation, challenging some traditional accounts (e.g., Nakayama, He, & Shimojo 1995). By Grill-Spector and Kanwisher's account, "as soon as you know it is there you know what it is."

Grill-Spector and Kanwisher (2005) conducted a series of experiments contrasting the time course of detecting the presence of any object, categorizing an object at the basic level, and categorizing an object at the subordinate level. The time course of object detection and basic-level categorization were equivalent and both were significantly faster than the time course of subordinate-level categorization. This suggests that at the same time during visual perception that an object is detected its basic-level category is known, but further processing is needed for subordinate-level categorization (but see discussion later, as well as Bowers & Jones, 2008; Mack, Gauthier, Sadr, & Palmeri, 2008).

Grill-Spector and Kanwisher acknowledged that an equivalent time course for object detection and basic-level categorization can be explained by two different accounts: Either object detection and basic-level categorization are performed by the very same mechanism or object detection and basic-level categorization are mediated by separate mechanisms that merely have a similar time course. To test between these competing accounts, they reported an experiment that directly assessed the dependence between object detection and basic-level categorization on a trial-by-trial basis. Their experiment required participants to make two responses on each trial, an object detection decision and a categorization decision. Each trial consisted of a briefly presented image followed by a mask followed by a second briefly presented image and then a final mask (see Figure 1). One of the two images was a nonobject texture pattern and the other image contained an object from one of two categories (person or car). Participants made an object detection response by deciding on which of the two images (the first or the second) contained an object and made an object categorization response by deciding whether that object was a person or a car. The dependence between object detection and object categorization was assessed by comparing the performance for one of the decisions conditionalized on the outcome of the other decision. According to their reasoning, if object detection and basic-level categorization are performed by the very same mechanism, then basic-level categorization should be successful only when object detection is successful, and vice versa; similarly, when object detection fails, basic-level categorization should fail as well, and vice versa. On the other hand, if basic-level categorization and object detection are not mediated by the same mech-

Michael L. Mack and Thomas J. Palmeri, Department of Psychology, Vanderbilt University.

This work was supported by a grant from the James S. McDonnell Foundation, NSF grant HSD-DHBS05, and the Temporal Dynamics of Learning Center (NSF Science of Learning Centers grant SBE-0542013).

Correspondences concerning this article should be addressed to Thomas J. Palmeri, Department of Psychology, Vanderbilt University, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817. E-mail: thomas.j.palmeri@vanderbilt.edu

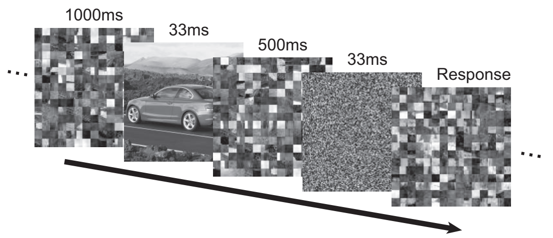

Figure 1. General trial sequence for all experiments reported in this article. Each trial began with a 1000 ms mask, followed by an image displayed for variable duration (Experiment 1: 17 or 33 ms; Experiment 2: 17, 33, 50, 68, or 167 ms; Experiments 3-5: 33 ms), then another mask displayed for 500 ms, followed by a second image displayed for the same duration as the first image, and then a final mask. One of the two images contained an object and the other image was a nonobject texture pattern. Responses could be made from the onset of the second image.

anisms, then the outcome of basic-level categorization should not depend on the outcome of object detection, and vice versa. Grill-Spector and Kanwisher's results were consistent with the existence of a single mechanism for detection and basic-level categorization. As a contrast, Grill-Spector and Kanwisher also had participants perform a block of trials with object detection and subordinate-level categorization (Harrison Ford vs. other man). In that case, no dependence was found; successful object detection did not depend on successful subordinate-level categorization.

Grill-Spector and Kanwisher provide a provocative description of visual object processing. After early visual processing, a stage of processing both parses the world into unique objects and categorizes those objects at the basic level. Finer-level categorization and coarser-level categorization occur only after this initial stage of detection and basic-level categorization has completed. By this account, knowing what basic-level category an object belongs to could possibly serve to shuttle visual information to the appropriate category-specific higher-level processing system.

But this theoretical claim must bear some scrutiny because it runs counter to a range of successful models of visual object categorization and identification (e.g., Joyce & Cottrell, 2004; Nosofsky & Kruschke, 1992; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; see also Palmeri et al., 2004). According to many models, explicit decisions about objects, whether categorization, identification, or detection, are not intimately intertwined within the hierarchy of visual processing per se. Some categorizations might be fast because those decisions are easy (Nosofsky & Palmeri, 1997) or they might be fast because they are based on relatively low-level perceptual features that are available early (Lamberts, 2000), but particular kinds of decisions are not tied to particular stages of visual processing. If detection and categorization truly show a tight temporal and processing linkage, then these extant models are falsified.

Because of the potential theoretical importance of the Grill-Spector and Kanwisher results, we sought empirical evidence that might decouple object detection from categorization. Their experiment assessing the coupling of categorization and detection involved a discrimination of people versus cars. People and cars could be said to belong to two different basic-level categories. However, the contrast between them spans two different *superordinate-level* categories, animal (people) versus vehicle (cars), which furthermore spans the fundamental ontological distinction between living and nonliving things. Experiments contrasting different levels of categorization need to take into account not only the within-category coherence, as in cars being a basic-level category, but also the between-category contrast participants are asked to make in the experiment, as in discriminating cars versus people (Mandler, Bauer, & McDonough, 1991). Otherwise, it is impossible to know whether performance is driven by the within-category coherence (at the basic level) or between-category contrast (at the superordinate level).

Bowers and Jones (2008) addressed the issue of between-category contrast in an experiment comparing response times for object detection and categorization with unmasked objects that were displayed briefly. Response times were equivalent for object detection and categorization defined by a superordinate-level contrast (e.g., dog vs. bus), but were significantly faster for detection than categorization defined by a basic-level contrast (e.g., dog vs. cat). This suggests that object detection can occur faster than basic-level categorization and that between-category contrast is one critical factor in defining categorization speed.

How might this impact the interpretation of the Grill-Spector and Kanwisher results? For example, it is possible that both detection and superordinate categorization might rely on relatively low-level visual features such as curvilinear versus rectilinear edges information (e.g., Levin, Takarae, Miner, & Keil, 2001). Or perhaps both decisions could be made based on rapid extraction of object parts. People have eyes. Cars don't. Cars have wheels. People don't. In both cases, participants could be discriminating cars versus people by essentially discriminating a sufficient amount of vehicle-like information from a sufficient amount of animal-like information. The coupling between detection and superordinate categorization (as opposed to basic-level categorization) might be caused by the dependency of both low-level features or rapidly identified parts, not a complete, categorized representation of whole objects.

A final factor that may underlie the coupling of object detection and categorization observed by Grill-Spector and Kanwisher is that one of the target categories was people and most of the stimuli they used in this category included full views of faces. The tight coupling they observed between detection and categorization could be explained by the special nature of faces (people) rather than some general mechanism underlying early stages of object recognition. To begin with, there are theoretical claims of an initial stage of processing that segments and categorizes objects as faces before their identification (Tsao & Livingstone, 2008; see also Mack et al., 2009) with support from time-course measures of neural activity (e.g., Anaki, Zio-Golumbic, & Bentin, 2007; Liu, Harris, & Kanwisher, 2002). Second, normally-functioning adults can be considered face experts (Carey, 1992; Diamond & Carey, 1986; Gauthier & Tarr, 1997; Tanaka, 2001). In the same way that bird experts identify bird images as quickly at a subordinate level as they are categorized as birds at the basic level (Tanaka & Taylor, 1991), familiar faces are identified as quickly as unique individuals as they are categorized as people (Tanaka, 2001). Both of these accounts suggest a priority for perception of faces that translates into faster and more accurate categorization of people relative to other kinds of (nonexpert) objects. Whatever the explanation, these alternate factors are very different from proposing that a relatively early stage of visual processing both detects the presence of objects (any object) and provides their basic-level categorization.

The present study directly addresses the finding by Grill-Spector and Kanwisher that success at basic-level categorization depends inextricably on success at object detection and vice versa. We report five experiments that all followed the experimental procedures used by Grill-Spector and Kanwisher (2005). Experiment 1 is a direct replication of Grill-Spector and Kanwisher's experiment 4 that also serves to explicate the methods and analyses used throughout the rest of this article.

Experiment 2 examines whether there is a behavioral dependence between object detection and categorization, irrespective of what categories are contrasted. If the contrasting category does not matter, then a tight coupling should be observed whether the discrimination is between cars and people, cars and boats, or cars and trucks. In all cases, the discrimination involves two basic-level categories. So the coupling between detection and categorization should be the same. This finding would bolster Grill-Spector and Kanwisher's theoretical claim and lead to a serious reexamination of most extant models of object recognition and categorization. On the other hand, if the coupling observed by Grill-Spector and Kanwisher is driven by the superordinate contrast between cars and people, not the basic-level status of the two categories considered independently, then the coupling should fall off as the between-category discrimination becomes more like a true basic-level discrimination (cars versus boats or cars versus trucks). Indeed, this is what we observe, indicating that object detection and basic-level categorization do not constitute components of the same early stage of visual processing.

In the remaining three experiments, we turn to an examination of what factors may underlie the tight coupling of detection and categorization that was observed by Grill-Spector and Kanwisher. First, we assess whether this coupling generalizes to superordinate-level contrasts that do not contain the fundamental distinction of living versus nonliving objects. Second, we explore the possibility of a priority of processing for people.

Experiment 1

Methods

Participants. Sixteen undergraduate students from Vanderbilt University participated in the experiment for course credit.

Stimuli. Stimuli were images of faces and cars selected from the same image database used by Grill-Spector and Kanwisher (2005). Nonobject textures were created by randomly scrambling 1×1 pixel squares from natural images and pattern masks were created by randomly scrambling 8×8 pixel squares from natural images. Stimuli subtended approximately $5.2° \times 5.2°$ of visual angle and were presented on a 19-in computer monitor that sat approximately 60 cm from the participants. No images were repeated during the experiment.

Procedure. The procedure was also identical to that used by Grill-Spector and Kanwisher (2005). Participants completed four blocks of trials, two comparing categorization (person vs. car) to object detection and two comparing identification (Harrison Ford vs. another man) to object detection. A block of trials began with instructions explaining the type of perceptual decision required during the block (categorization or identification). Each trial consisted of an image shown for 17 or 33 ms, followed by a mask shown for 500 ms, a second image shown for 17 or 33 ms, and a final mask shown until 3 s had elapsed from the onset of the trial. One of the images contained an object and the other image a nonobject texture. For the categorization blocks, half of the trials showed an image of a person and remaining trials a car; for the identification blocks, half of the trials showed an image of Harrison Ford and remaining trials a random nonfamous male person. With the onset of the second image, participants were asked to make two perceptual decisions. Participants responded to the object detection decision by identifying which image contained the object. Responses were made by pressing the "1" (first image) or "2" (second image) key. Participants responded to the category or identity of the object image. This decision was between a person or car in the categorization block and between Harrison Ford or other man in the identification block. Responses were made by pressing one of two keys labeled with choices for the particular block. If 3 s elapsed before both responses were made, participants were warned to respond faster, the trial was flagged for removal, and the next trial began. The experiment consisted of four blocks combining exposure duration (17 or 33 ms) and categorization versus identification decision. Each block consisted of 128 trials. The order of blocks was randomized across participants. The order of responses to detection or to categorization/identification was counterbalanced across participants. The entire experiment lasted approximately 45 min.

Results

Categorization and identification blocks were analyzed separately and in the same manner. Data from the two response orders were averaged together. Since the logic can be a bit complicated, we will first highlight the analysis steps leading to the data presented in Figure 2 and what they might indicate about a coupling between detection and categorization.

In order to assess the dependence between object detection and categorization, analyses conditionalized performance for one decision based on the outcome of the other decision. Specifically, the probability that categorization (gray bars) was successful – P(hit) – was calculated for trials when object detection was also successful (*hit* on the x axis) and when object detection failed (*miss* on the x axis). Similarly, the probability that object detection (black bars) was successful – P(hit) – was calculated for trials when categorization was also successful (*hit* on the x axis) and when categorization failed (*miss* on the x axis).

According to Grill-Spector and Kanwisher (2005), if categorization and object detection are tightly coupled, then categorization performance should be good when detection is a hit but at chance when detection is a miss; similarly, detection performance should be good when categorization is a hit but at chance when categorization is a miss. Specifically, in terms of the experimental factors in the analyses, if detection and categorization are tightly coupled, there should be a significant main effect of Success on the other task (better performance for a hit on the other task than a miss) but no main effect of Task (detection or categorization) and no interaction between the two factors. Furthermore, performance for both tasks should be at chance when the other task is a miss. That is essentially what Grill-Spector and Kanwisher found.

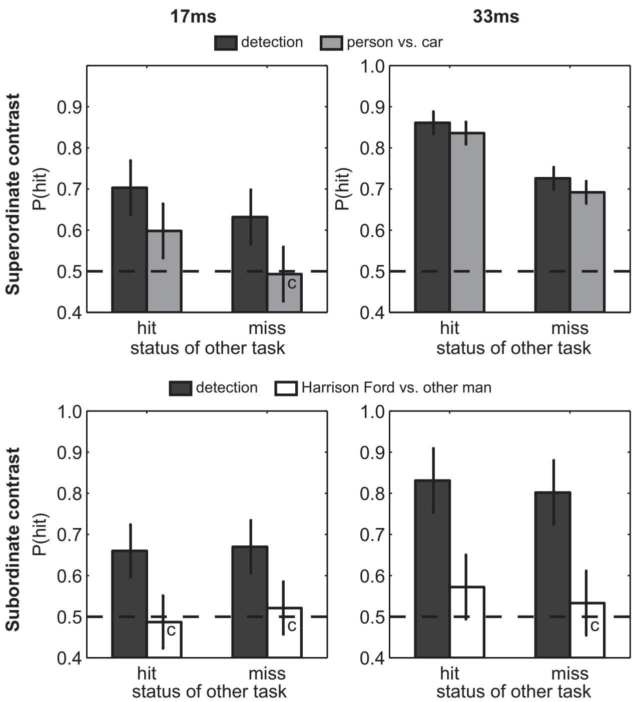

Figure 2. Results from Experiment 1. The top row of panels shows probability of a hit for object detection (black bars) and categorization defined by a superordinate contrast (dark gray bars) depending on whether the outcome of the other decision was a success (hit) or not (miss). The bottom row of panels shows probability of a hit for object detection (black bars) and identification (white bars) depending on whether the outcome of the other decision was a success (hit) or not (miss). The left column of panels show results from trials with exposure durations of 17 ms and the right column of panels show results from trials with exposure durations of 33 ms. Chance level (50%) is indicated by the black dashed line and conditions with performance equivalent to chance levels are indicated by a black letter "c" on the bar. Error bars in this figure and the remaining figures represent confidence intervals of the Task main effect with α level of 0.05.

To maintain consistency with the analyses they reported, we report separate analyses for the two different exposure durations (17 ms and 33 ms) and for categorization/detection and identification/detection blocks. For the 17 ms exposure condition, we conducted a 2 (Task – detection versus categorization) × 2 (Success – hit versus miss on the other task) analysis of variance (α level of 0.05 and Greenhouse-Geisser corrected) to assess the dependence between the two decisions. Details of this analysis are presented in Table 1. Both categorization and detection were significantly better when the other task was a hit than when the other task was a miss, as reflected by a significant main effect of Success; object detection performance is good when categorization is good and vice versa. Performance for object detection was better than for categorization as revealed by a significant main effect of Task. The interaction of Task × Success was not significant; detection and categorization were affected by the success of the other task in the same manner. Planned comparisons were also conducted to assess whether performance was at chance levels (denoted in the figures as a "c" on the bar). Performance was at chance for categorization when the other decision was a miss.

For categorization/detection trials with 33 ms exposure duration (upper right panel of Figure 2 and Table 1), there was a significant main effect of Success of the other task, a marginally significant main effect of Task, but no interaction observed. Planned comparisons revealed that performance was above chance for all conditions. Similarly, Grill-Spector and Kanwisher (2005) observed chance level categorization performance only in the 17 ms exposure condition.

Like Grill-Spector and Kanwisher, the identification/detection trials results were quite different (lower panels of Figure 2 and Table 1). Object detection was significantly better than identification, as revealed by a significant main effect of Task for both the 17 ms and 33 ms exposure durations. No significant main effect of Success or an interaction was observed.

Discussion

The results of Experiment 1 weakly replicate Grill-Spector and Kanwisher (2005). When asked to detect the presence of an object and categorize that object as a person versus a car, participants showed some evidence of dependence between the two decisions. With 33 ms exposure duration, the success of the other task influenced performance on the current task, there was only a marginal difference between detection and categorization performance, and no interaction between the two factors. In other words, the success of categorization depended on the success of detection and vice-versa. On the other hand, when asked to detect an object and identify it uniquely (Harrison Ford vs. other man), participants showed no dependence. One important piece of evidence for dependence is that both categorization and object detection should be at chance when the other decision is a miss. We only observed this for categorization performance in the 17 ms exposure duration condition, which somewhat weakens the strong claim for a tight coupling between object detection and categorization. Similarly, the difference in performance for object detection and categorization observed in the 17 ms exposure duration condition does not support a dependency between these two decisions. Successful object detection did not necessarily lead to successful categorization. This failure to fully replicate Grill-Spector and Kanwisher's results may arise from an issue of statistical power. The same effects we observed may be present in their results, yet their smaller sample size ($N = 12$) may have led to an underpowered analysis.

Like Grill-Spector and Kanwisher, this experiment used a superordinate contrast between people and cars. The next two experiments explored a basic-level contrast instead.

Experiment 2

If object detection and basic-level categorization are tightly coupled, then the nature of the between-category contrast should be irrelevant to performance. Whether the task asks participants to categorize an object as a car versus a person, or categorize an object as a car versus a boat, the results should be the same. Both contrasts use basic-level categories. Success at categorization should depend on success at object detection and vice versa, as we observed in Experiment 1 and as observed by Grill-Spector and Kanwisher (2005). On the other hand, if the between-category contrast matters, then little coupling between detection and cate-

Table 1
Analysis of Variance for Experiments 1 and 2

	Contrast	Exposure duration	Factor	F	MSE	p	η_p^2
Experiment 1	Superordinate (person vs. car)	17	Task	6.959	.031	**.019**	.317
			Success	10.406	.011	**.006**	.410
			T × S	2.476	.002	.136	.142
		33	Task	3.717	.005	.073	.199
			Success	17.315	.017	**.001**	.536
			T × S	1.175	.001	.296	.073
	Identification (Harrison Ford vs. other man)	17	Task	34.866	.010	**.001**	.699
			Success	1.519	.007	.237	.092
			T × S	1.287	.001	.274	.079
		33	Task	94.361	.012	**.001**	.863
			Success	.566	.018	.463	.036
			T × S	.003	.004	.959	.001
Experiment 2	Basic (dog vs. bird, car vs. plane, chair vs. bed)	17	Task	26.373	.013	**.001**	.706
			Success	1.241	.014	.289	.101
			T × S	2.331	.001	.155	.175
		33	Task	18.997	.021	**.001**	.633
			Success	16.719	.017	**.002**	.603
			T × S	3.699	.012	.081	.252
		50	Task	44.508	.008	**.001**	.802
			Success	17.382	.016	**.002**	.612
			T × S	11.256	.005	**.006**	.506

Note. Contrast describes the level of categorization performed along with object detection; the actual contrast is shown in parentheses. Exposure duration is in milliseconds. ANOVAs were conducted with $\alpha = 0.05$ and Greenhouse-Geisser corrected.

gorization should be observed. In Experiment 2, we compared performance for object detection to categorization of basic-level categories separated by only a basic-level contrast. We also further examined the role of perceptual encoding on object detection and categorization. Coupling between detection and basic-level categorization may be evident with shorter or longer time for encoding the stimuli. We tested the role of encoding by systematically varying exposure durations of the test stimuli.

Methods

Participants. Twelve Vanderbilt University students participated in this experiment for course credit.

Stimuli. Stimuli were images of dogs, cars, planes, boats, chairs, and beds. We used the same images of cars from Experiment 1. Stimuli for the other categories were selected from the same database used by Grill-Spector and Kanwisher (2005) and various web sources. Stimulus presentation was the same as Experiment 1.

Procedure. This experiment followed the same procedure as outlined in Experiment 1 with the following exceptions: Experiment 2 used five exposure durations (17, 33, 50, 68, and 167 ms) to more fully map out the effect of encoding time on detection and categorization. Participants completed two blocks of each basic-level contrast (dog vs. bird, car vs. plane, or chair vs. bed). The entire experiment consisted of 540 trials and lasted approximately 45 min. The order of the blocks was randomized across participants and exposure duration order was randomized within a block.

Results

There were no qualitative differences between the three versions of basic-level contrast (dog vs. bird, car vs. plane, or chair vs. bed), so data were collapsed across the contrasts for analyses. Participants committed few errors with exposure durations of 68 and 167 ms thereby preventing conditional analyses from being reported. Thus, we first analyze conditional performance for exposure durations of 17, 33, and 50 ms, as shown in Figure 3.

Conditional performance for exposure durations of 17, 33, and 50 ms show a marked difference from Experiment 1. While performance was better on object detection and categorization when the other decision was a success, performance for object detection was significantly better than categorization whether the categorization was a success and not. Analysis of variance was conducted as in Experiment 1 for each of the three exposure durations (see Table 1 for analysis of variance [ANOVA] details). For the 17 ms exposure condition, detection performance was better than categorization performance as revealed by a significant main effect of Task. Neither the main effect of Success nor the interaction was significant. For the 33 and 50 ms exposure conditions, significant main effects of Task and Success were observed. Performance was better for both detection and categorization when the other task was a success but performance for detection was always better than categorization. The Task × Success interaction was also significant at both exposure durations. The difference between detection and categorization performance was significantly smaller when the other task was successful than when it was not.

The time course of overall accuracy for object detection (dark line) and categorization (gray line) with a basic-level contrast is shown in Figure 4. We observed better performance for detection than categorization at short exposure durations; for long exposure durations, performance was at ceiling. A 2 (Task) × 5 (Exposure Duration) analysis of variance revealed significant main effects of Task [$F(1, 11) = 56.6$, $MSE = 0.003$, $\eta_p^2 = 0.837$] and Exposure Duration [$F(1, 11) = 139.5$, $MSE = 0.008$, $\eta_p^2 = 0.927$]; perfor-

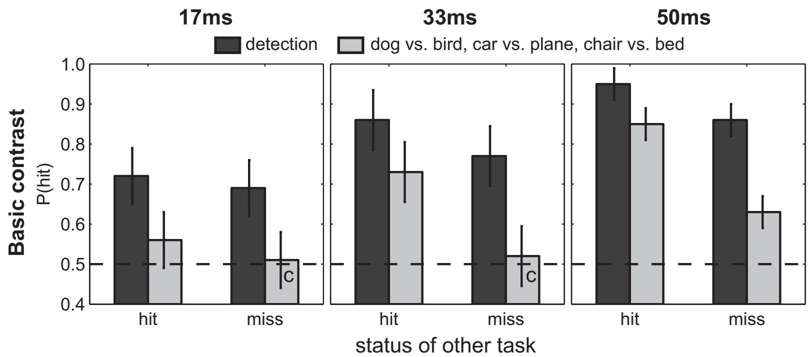

Figure 3. Results from Experiment 2. The three panels show probability of a hit for object detection (black bars) and categorization defined by a basic-level contrast (light gray bars) dependent on the outcome of the other task (hit or miss) for exposure durations of 17 (left), 33 (middle), and 50 ms (right).

mance was better for detection than categorization and performance was better with longer exposure durations. A significant interaction of Task × Exposure Duration [$F(1, 11) = 14.12$, $MSE = 0.005$, $\eta_p^2 = 0.562$] was also observed; performance was better for detection than categorization at short exposure durations, but was equivalent (at ceiling) at longer exposure durations. Planned comparisons at each exposure duration were consistent with the analysis of variance with significant differences for 17-50 ms exposure [t's$(11) > 5$] but not for 68 or 167 ms exposure, [t's$(11) < 1$].

Discussion

Observing a coupling between categorization performance and object detection performance depends critically on the between-category contrast. Unlike contrasts defined at a superordinate level, for contrasts defined at a basic level (dog vs. bird, car vs. plane, or chair vs. bed), there was no tight coupling of object detection and categorization for 17, 33, and 50 ms exposure durations. Detection was better than basic-level categorization. While basic-level categorization was nearly at chance levels when object detection failed, object detection was often successful when categorization failed. Participants could detect an object was there without being able to tell whether it was a dog or a bird (or a car or a plane, or a chair or a bed).

Performance on detection and categorization was affected to some degree by the success of the other task; categorization accuracy was higher when detection was a success and vice versa. Detection and categorization may not be independent, which is perhaps not surprising given that they are based on much of the same perceptual information. The results of the conditional analysis from Experiment 2 clearly do not support a claim of strong dependence between object detection and basic-level categorization as the same process with the same temporal dynamics.

Our observed decoupling of detection and basic-level categorization was evident over a range of exposure durations and different object categories. With shorter encoding time, object detection is significantly more successful than categorization, but this advantage is eliminated with longer encoding times only because performance reached ceiling. The decoupling revealed in this experiment is inconsistent with notion that an initial stage of processing that both detects objects and accesses their basic-level category before further visual processing.

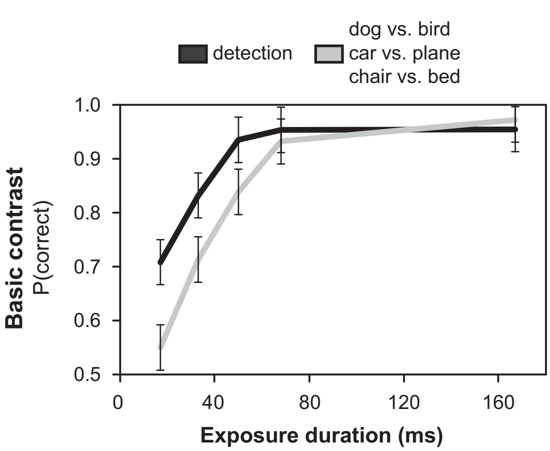

Figure 4. Overall accuracy results from Experiment 2. The graph plots percent correct for object detection (black line) and categorization defined by a basic-level contrast (light gray line).

Experiment 3

Both Grill-Spector and Kanwisher (2005) and our Experiment 1 do suggest that certain factors may lead to some coupling between object detection and categorization. The next three experiments investigated the nature and extent of this coupling.

As noted in the introduction, the type of categorization contrast used by Grill-Spector and Kanwisher (person vs. car) can be characterized by three factors: a superordinate-level contrast (animal vs. vehicle), a fundamental distinction between living and nonliving objects, and the inclusion of the socially-relevant, expert category of people and faces. In the following experiment, we investigate the dependence between detection and categorizations defined by a superordinate-level contrast that were not confounded by including people or by including contrasts that cross the living/nonliving distinction.

Methods

Participants. Fifty Vanderbilt University undergraduate students participated in the experiment for course credit.

Stimuli. The stimuli consisted of images of birds, dogs, flowers, trees, beds, chairs, cars, and boats. Some images were the same as used in Experiments 1 and 2; images of flowers and trees of similar size and image properties to those were gathered through various web sources. Image presentation was the same as Experiment 1. No image was repeated during the experiment.

Procedure. This experiment followed the same procedure as outlined in Experiment 1 with the following exceptions: Participants were assigned to one of two groups: one that was presented with only living objects and another that was presented with only nonliving objects. In this way, no participant was tested on a contrast that crossed the living/non-living boundary. Each participant completed four blocks of trials: two blocks of trials with object detection and a superordinate-level contrast categorization and two blocks of trials with object detection and a basic-level contrast categorization. Object categories were chosen so that each category was part of a superordinate-level as well as a basic-level contrast (Living/Superordinate: bird vs. flower, tree vs. dog; Living/Basic: birds vs. dog, flower vs. tree; Nonliving/Superordinate: bed vs. car, boat vs. chair; Nonliving/Basic: bed vs. chair, boat vs. car). Stimuli were presented for a fixed exposure duration (33 ms). The experiment consisted of 512 trials and lasted approximately 45 min.

Results

All analyses were performed in the same manner as Experiment 1. There were no qualitative differences across objects within the conditions, so data were collapsed within each unique condition. Results for Experiment 3 are summarized in Figure 5 and analysis of variance results are presented in Table 2.

For both living and nonliving object groups and for both superordinate- and basic-level contrasts, trials comparing detection to categorization showed main effects of Task and Success and a significant interaction (see Table 2). Overall performance was better when the other task was successful. However, detection performance was better than categorization performance and the difference between detection and categorization was larger when the other task was a miss. Performance was above chance for all conditions.

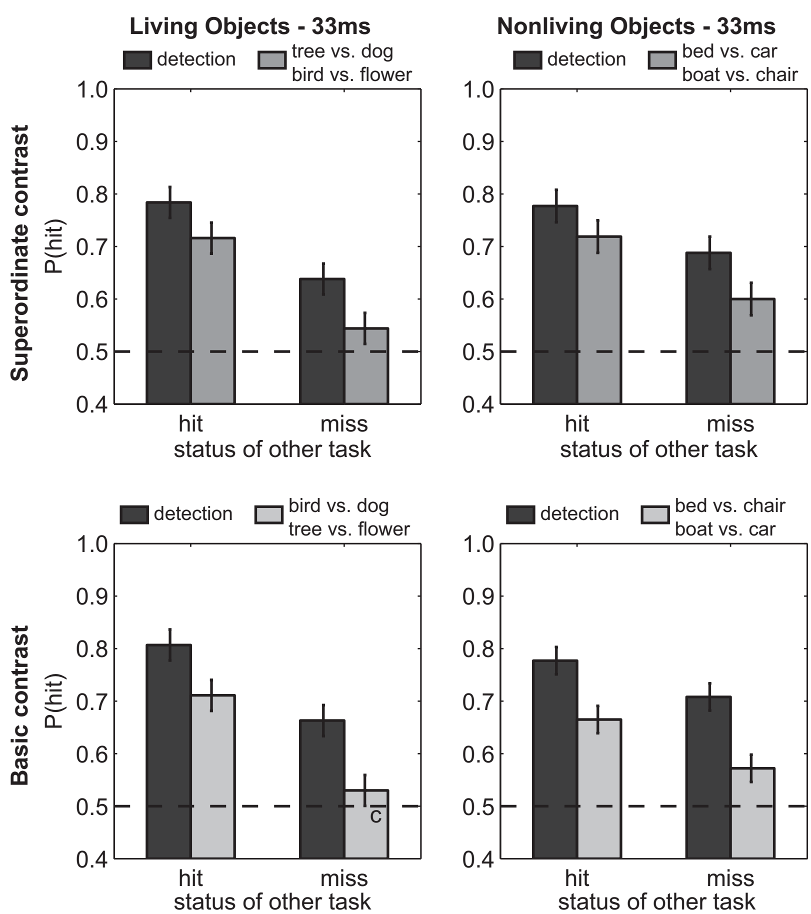

Figure 5. Results from Experiment 3. The four panels show probability of a hit for object detection (black bars) and different types of categorization depending on the outcome of the other task (hit or miss). Performance for superordinate contrasts is shown in the top row, basic-level contrasts in the bottom panels, living objects in the left column, and nonliving objects in the right column.

Table 2
Analysis of Variance for Experiments 3, 4, and 5

	Contrast	Exposure duration	Factor	*F*	*MSE*	*p*	η_p^2
Experiment 3	Superordinate (tree vs. dog, bird vs. flower)	33	Task	27.384	.006	**.001**	.533
			Success	59.633	.011	**.001**	.713
			T × S	30.997	.001	**.001**	.564
	Superordinate (bed vs. car, boat vs. chair)	33	Task	10.151	.013	**.004**	.297
			Success	34.755	.008	**.001**	.592
			T × S	6.572	.001	**.017**	.215
	Basic (bird vs. dog, tree vs. flower)	33	Task	55.181	.006	**.001**	.697
			Success	63.561	.010	**.001**	.726
			T × S	11.488	.001	**.002**	.324
	Basic (bed vs. chair, boat vs. car)	33	Task	11.490	.015	**.002**	.324
			Success	35.029	.011	**.001**	.593
			T × S	5.666	.001	**.026**	.191
Experiment 4	Superordinate (person vs. car)	33	Task	.016	.009	.901	.001
			Success	22.71	.018	**.001**	.497
			T × S	1.101	.001	.305	.046
	Superordinate (dog vs. boat)	33	Task	7.279	.011	**.013**	.240
			Success	27.920	.010	**<.001**	.548
			T × S	11.751	.001	**.002**	.338
	Basic (person vs. dog)	33	Task	11.97	.017	**.002**	.342
			Success	74.62	.010	**.001**	.764
			T × S	10.23	.001	**.004**	.308
	Basic (car vs. boat)	33	Task	15.360	.017	**.001**	.401
			Success	86.472	.007	**<.001**	.784
			T × S	23.853	.001	**.002**	.509
Experiment 5	Superordinate (person vs. flower, person vs. tree)	33	Task	.102	.022	.752	.003
			Success	35.264	.030	**<.001**	.569
			T × S	1.31	.004	.263	.014
	Superordinate (person vs. boat, person vs. chair)	33	Task	3.212	.018	.085	.084
			Success	38.02	.032	**<.001**	.591
			T × S	12.64	.002	**.002**	.378
	Basic (person vs. bird, person vs. cat)	33	Task	9.292	.015	**.005**	.243
			Success	31.450	.029	**<.001**	.555
			T × S	2.751	.002	.110	.074

Note. Contrast describes the level of categorization performed along with object detection; the actual contrast is shown in parentheses. Exposure duration is in milliseconds. ANOVAs were conducted with $\alpha = 0.05$ and Greenhouse-Geisser corrected.

Discussion

We found no evidence that categorization defined by a superordinate-level contrast is strongly coupled with object detection. Performance on superordinate-level contrast categorization and detection was decoupled across several versions of categorization between objects within the categories of living things and nonliving things. Performance in both detection and categorization was affected by the success of the other task, so the tasks are not independent. However, it was often the case that objects were detected without being categorized. Similar to Experiment 2, we found further evidence that performance for categorization defined by basic-level contrasts was dissociated from object detection. These results suggest that the dependence between categorization and detection observed by Grill-Spector and Kanwisher does not generalize to other categorizations defined by a superordinate-level contrast.

Experiment 4

We next investigated more closely the effect of people as a target category on detection and categorization. People may have priority in processing, whether that is because of some initial stage that categorizes faces as faces (e.g., Anaki et al., 2007; Liu et al., 2002) or because of face expertise (e.g., Diamond & Carey, 1984; Gauthier & Tarr, 1997; Tanaka, 2001). Whatever the reason, that priority could lead to person categorization performance being more closely tied to object detection, at least temporally if not also mechanistically.

Methods

Participants. Twenty-five Vanderbilt University students participated in this experiment for course credit.

Stimuli. Stimuli were images of people, dogs, cars, and boats. We used the same images of people and cars from Experiment 1. We added images of dogs and boats that had similar size and image properties. Dogs and boats were included to create basic-level contrasts with the categories used in Experiment 1: people vs. dogs and cars vs. boats are basic-level contrasts within the same superordinate-level categories. No images were repeated during the experiment. Stimulus presentation was the same as Experiment 1.

Procedure. This experiment followed the same procedure as outlined in Experiment 3 with the following exceptions: Participants completed two blocks with an object detection decision and a categorization defined by a superordinate-level contrast (person vs. car or dog vs. boat) and two blocks with an object detection and a categorization defined by a basic-level contrast (person vs. dog or car vs. boat). Participants started with either the two superordinate-level or two basic-level contrast blocks and the order of the two blocks within a contrast level was randomized. The entire experiment consisted of 512 trials and lasted approximately 45 min.

Results

Each block was analyzed separately and in the same manner as in the previous experiments. We will refer to each of these blocks by the categorization performed during the block (person vs. car, dog vs. boat, person vs. dog, and car vs. boat). Results for Experiment 4 are shown in Figure 6 and analysis of variance results are presented in Table 2.

We begin discussing the two versions of the superordinate contrast blocks. For the person vs. car contrast condition (top left panel of Figure 6), both categorization and detection were significantly better when the other task was a hit than when the other task was a miss, as reflected by a significant main effect of Success. Neither the main effect of Task nor the interaction of Task × Success were significant.

For the other superordinate-level contrast condition, dog vs. boat (top right panel of Figure 6 and Table 2), there was also a significant main effect of Success of the other task; performance on both categorization and detection was better when the other task was successful. However, in contrast to the superordinate contrast of person vs. car, a significant main effect of Task and a significant interaction were also observed. Detection performance was significantly better than categorization with a larger difference between the two tasks when the other task was a miss than a hit.

For the two contrasts at the basic level (bottom panels of Figure 6 and Table 2), there was also a significant main effect of Success of the other task; categorization defined by basic-level contrasts as well as detection were better when the other task was a hit than when the other task was a miss. Like the superordinate dog vs. boat condition, but in contrast to the superordinate person vs. car condition, a significant main effect of Task and a significant interaction were observed. Performance was significantly better for detection than categorization with a greater difference when the other task was a miss than a hit.

Discussion

When categorizing a person versus a car, a superordinate-level contrast, there is a coupling in performance for object detection and categorization like that found by Grill-Spector and Kanwisher (2005). However, this strong dependence does not generalize to

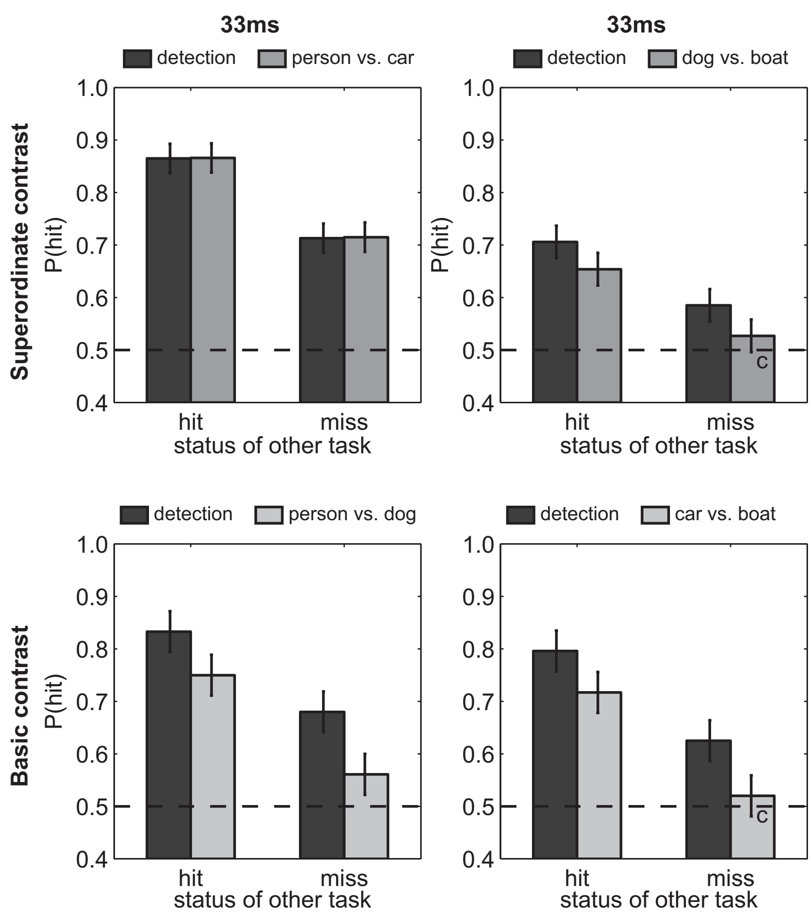

Figure 6. Results from Experiment 4. Probability of a hit for object detection (black bars) and categorization defined by a superordinate contrast (dark gray bars) is shown in the top row and categorization defined by a basic-level contrast (light gray bars) is shown in the bottom row.

categorizing dog versus boat, a superordinate-level contrast with a similar ontological span. As we found in Experiments 2 and 3, object detection is decoupled from categorization of a basic-level contrast even when one of the target categories is people. These result perhaps suggests some priority in processing for categorization of people, but only when the contrasting category is sufficiently dissimilar to people.

Experiment 5

Finally, Experiment 5 attempted to generalize the results of Experiment 4 to an extended set of categorizations involving people. We compared object detection and categorization for a superordinate-level contrast of people versus other living objects, a superordinate-level contrast of people versus nonliving objects, and a more basic-level contrast with people.

Method

Participants. Twenty-six Vanderbilt University undergraduate students participated in the experiment for course credit.

Stimuli. Stimuli consisted of images of people, birds, cats, boats, chairs, flowers, and trees. Image presentation was the same as in Experiment 1 and no images were repeated during the experiment.

Procedure. The procedure of this experiment were the same as Experiment 3 with the following exceptions: Participants completed six blocks of trials: two blocks of object detection and categorization defined by a superordinate-level contrast of living objects (person vs. flower, person vs. tree), two blocks of object detection and categorization defined by a superordinate-level contrast of nonliving objects (person vs. boat, person vs. chair), and two blocks of object detection and categorization defined by a more basic-level contrast (person vs. bird, person vs. cat). Participants completed the two blocks of each type of between-category contrast one after the other and the order of the pairs of blocks was randomized across participants. The experiment consisted of 512 trials and lasted approximately 45 min.

Results

All analyses were performed in the same manner as the previous experiments. There were no qualitative differences between the two blocks of superordinate-level contrasts with living object or nonliving object or basic-level contrasts, so data were collapsed across the versions for analyses. Results for Experiment 5 are shown in Figure 7 and detailed analysis of variance results are presented in Table 2.

For trials comparing object detection to categorization of people in superordinate-level contrasts, we largely replicated the findings of Grill-Spector and Kanwisher and our Experiment 1. For the condition with people versus other living objects, a significant main effect of Success was observed, with no significant effect of Task or interaction. For the condition of people versus nonliving objects, a significant main effect of Success and significant interaction was observed with no main effect of Task. The performance of both object detection and categorization was affected by the success of the other task. By contrast, for the more basic-level contrasts, results largely replicated those observed in Experiments 2–4. Analysis revealed significant main effects of Task and Success but no significant interaction.

Discussion

Categorizing people versus a sufficiently dissimilar category results in a performance that is tightly coupled with object detection, whether the superordinate-level contrast was between people and other living objects (people versus flowers or trees) or between people and nonliving objects (people versus boats or chairs). However, as we observed throughout this article, categorization defined by a more basic-level contrast, regardless of the target categories, shows a decoupling of categorization from object detection.

General Discussion

In a series of experiments, Grill-Spector and Kanwisher (2005) reported an identical time course for object detection and basic-level categorization and reported a strong behavioral coupling between the success of object detection and the success of basic-level categorization. These results led to a provocative suggestion that object detection and basic-level categorization may be the very same mechanism. During relatively early stages of visual processing, objects are extracted from the background through figure-ground segregation mechanisms *and* those objects are categorized at the basic level. It is possible that this basic-level categorization helps shuttle visual processing to the appropriate higher-level category-specific processing areas; a similar theoretical claim has been recently proposed as a specialized mechanism for detecting and segmenting faces from the background (Tsao & Livingstone, 2008). While intriguing, this theoretical suggestion does run counter to most extant models of visual object categorization and many models of face recognition. If object detection and basic-level categorization are tightly coupled, or if they are the same mechanism, many current models are wrong.

Mack et al. (2008) recently addressed the first part of Grill-Spector and Kanwisher's evidence. If object detection and basic-level categorization are both performed by the same mechanism at the same early stage of visual processing, then any manipulation that affects basic-level categorization should similarly affect object detection. Mack et al. decoupled the time course of object detection and basic-level categorization by introducing stimulus manipulations like image inversion and image degradation. Categorization was impaired by image inversion but object detection was not. Image degradation had a significantly greater effect on basic-level categorization than object detection. Decoupling object detection and basic-level categorization argues against a common mechanism that does both. However, both of these manipulations involved an explicit image manipulation. As acknowledged by Grill-Spector and Kanwisher, an extreme amount of image degradation would dissociate object detection and categorization for rather trivial reasons: We can detect a large bird in the air from a quarter mile away, but may be unable to tell whether it's a bird, a plane, or Superman.

The current experiments used no image degradation, but instead manipulated the contrast category during the task (Bowers & Jones, 2008; D'Lauro, Tanaka, & Curran, 2008; Mandler et al., 1991). The last experiment of Grill-Spector and Kanwisher asked

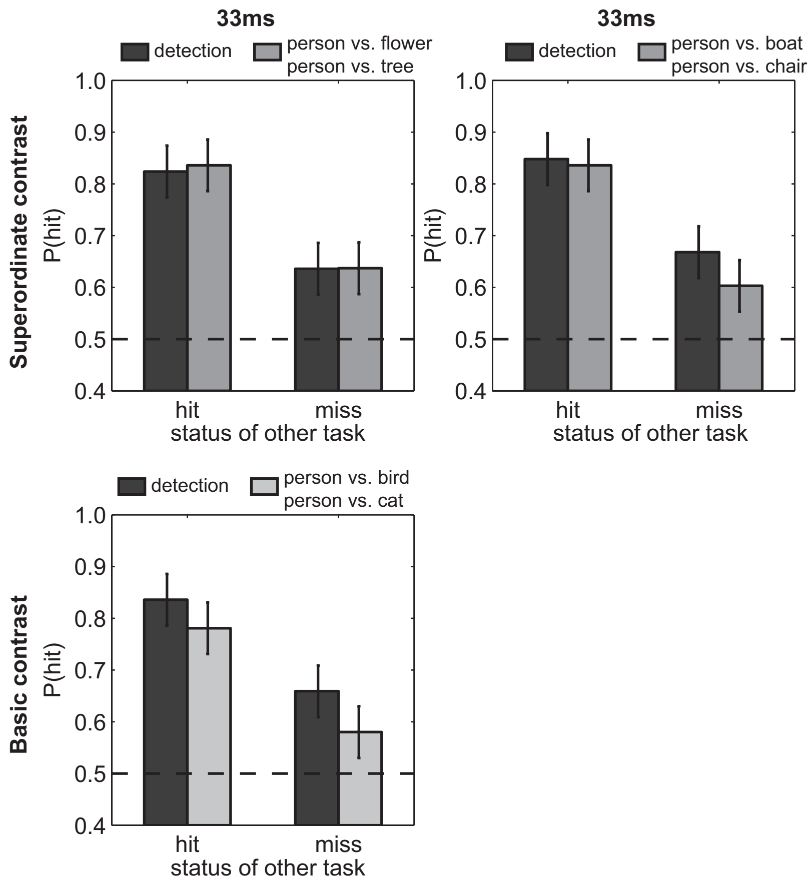

Figure 7. Results from Experiment 5. The three panels show probability of a hit for object detection (black bars) and different types of categorizations involving people. Top row shows performance for superordinate contrasts between people and living objects (left) and nonliving objects (right); bottom panel shows performance for basic-level contrast between people and living objects.

participants to detect which of two images contained an object and categorize that object as a car or a person. Success at detecting an object depended on success at categorizing that object and vice versa. Now while cars and people are arguably basic-level categories, the contrast the participants were asked to make spanned the superordinate level. If object detection and basic-level categorization are the same mechanism, then a tight behavioral coupling should be observed whether the contrast is between cars and people or between cars and boats. If detection and categorization are not linked, then manipulations of the between-category contrast may decouple performance between detection and categorization. Indeed, Bowers and Jones (2008) found a difference in mean reaction times for speeded object detection and basic-level categorization. The present experiments go beyond these studies by directly addressing whether *success* at basic-level categorization depends inextricably on success at object detection and vice versa.

We observed that success at object detection did not strongly depend on success at categorization when a more basic-level contrast was used (people vs. dogs or cars vs. boats). These results, combined with those of Mack et al. (2008) and Bowers and Jones (2008), are inconsistent with a tight behavioral coupling as well as a tight mechanistic coupling between object detection and *basic-level* categorization. Object detection is often faster than basic-level categorization. And the success of object detection does not depend on success of basic-level categorization. Object detection and basic-level categorization do not occur at the same early stage of visual processing.

With a superordinate-level contrast, the relationship between success for object detection and success for categorization depended on the object categories we used. In particular, we only observed a tight behavioral coupling when people were one of the relevant categories. All other versions of superordinate-level contrasts we tested showed an advantage for detection over categorization just like we observed for basic-level contrasts: Object detection was often more accurate than categorization and success of object detection did not depend on the success of a superordinate-level contrast with non-people objects. In other words, we observed a tight behavioral coupling of object detection and categorization only for categories used by Grill-Spector and Kanwisher—namely, superordinate-level contrast including people.

Overall, it is important to note that we observed very little evidence for a tight coupling between detection and categorization in any of the contrasts that we tested. Even with superordinate-level contrasts involving people, the strong claim of tight coupling was not supported by above chance performance in the miss condition. We typically found an interaction of the Task and Success factors where the difference between detection and categorization was larger when the other task was unsuccessful, an effect not supported by the predictions of either tight coupling or complete independence of detection and categorization. Rather, our results suggest at most a limited contingency between the tasks such that success on detection is correlated with better performance on categorization.

One interpretation of these results is that object detection and categorization become more behaviorally coupled for easier category contrasts. In a sense, an object detection decision is about the most superordinate categorization decision someone can make visually. In the present experiments, it's either a coherent, recognizable object or it's a pattern mask. Deciding whether something is a person versus a car is as easy as deciding whether something is an object or not. When there is sufficient visual information with rapid exposure to detect it, you can categorize it as a face or a car as well. This would not imply a stage of processing that both detects and categorizes, just that both the speed and success of detection and certain high-level categorizations relies on the same visual information that is available at the same time (see Lamberts, 2000). This description would be consistent with many extant theories of visual object categorization (e.g., Joyce & Cottrell, 2004; Lamberts, 2000; Nosofsky & Kruschke, 1992; Nosofsky & Palmeri, 1997; Palmeri, 1997; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; see also Palmeri et al., 2004). Some perceptual decisions can be made more quickly than others, but this does not require that faster decisions are tied to stages of visual processing that must take place before other decisions can be made.

Of course that still begs the question of whether there might be something special about people. To be clear, we did not find a tight behavioral coupling for every condition that involved people. For example, with people versus dogs, we did not. It was only with superordinate contrasts that included people, whether those contrasts cross the living/nonliving boundary, that we found this coupling. Certainly one possibility is that people (faces), as expert domains of processing, get a processing boost because of neural specialization or enhanced perceptual representations. That might serve to separate people versus cars or people versus trees even more than might be possible based solely on overall similarity or distance in an ontological hierarchy.

Another intriguing possibility is people do have some special place in the hierarchy of visual processing stages. Indeed, some theoretical claims posit an initial face segmentation stage (Tsao & Livingstone, 2008), there are time-course measures of neural activity that suggest a priority for processing faces (e.g., Anaki et al., 2007; Liu et al., 2002), and some behavioral work suggests a temporal advantage for categorizing people as people (Mack et al., 2009) that may generalize to other domains of expertise (Curby & Gauthier, 2009). There may not be an explicit stage that both detects and categorizes people or faces. But there could be relatively early representations that can be used for detecting properties of face-like or people-like stimuli in order to direct attention or to engage circuits, like the amygdala, involved in processing socially-relevant stimuli (e.g., Adolphs, 2009).

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60,* 693–716.

Anaki, D., Zio-Golumbic, E., & Bentin, S. (2007). Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *Neuroimage, 37,* 1407–1416.

Bowers, J., & Jones, K. (2008). Detecting objects is easier than categorizing them. *The Quarterly Journal of Experimental Psychology, 61,* 552–557.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions Royal Society London, B, 335,* 95–102.

Curby, K. M., & Gauthier, I. (2009). The temporal advantage for individuating objects of expertise: Perceptual expertise is an early riser. *Journal of Vision, 9*(7), 1–13, http://journalofvision.org/9/6/7/, doi:10.1167/9.6.7

D'Lauro, C., Tanaka, J., & Curran, T. (2008). The preferred level of face categorization depends on discriminability. *Psychonomic Bulletin & Review, 15,* 623–629.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673–1682.

Grill Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243–275.

Joyce, C., & Cottrell, G. W. (2004). Solving the visual expertise mystery. In H. Bowman & C. Labiouse. (Eds.), *Connectionist models of cognition and perception II: Proceedings of the eighth neural computation and psychology workshop.* Hackensack, NJ: World Scientific.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization, *Psychological Review, 107,* 227–260.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics, 63,* 676–697.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5,* 910–916.

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review, 15,* 28–35.

Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2007). Unraveling the time course of perceptual categorization: Does fastest mean first? *Proceedings of the Twenty-Ninth Annual Meeting of the Cognitive Science Society,* 1253–1258.

Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2009). Time course of visual object categorization: Fastest does not necessarily mean first. *Vision Research, 49,* 1961–1968.

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23,* 263–298.

Nakayama, K., He, Z. J., & Shimojo, S. (Eds.). (1995). *Visual surface representation: A critical link between lower-level and higher-level vision*: Cambridge, MA: The MIT Press.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation, 28,* 207–250. San Diego: Academic Press.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Science, 8,* 378–86.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Science, 104,* 6424–6429.

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534–543.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(4A), 57–482.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31,* 411–437.

Received September 22, 2008
Revision received April 6, 2010
Accepted April 13, 2010 ■